Journal of Applied Research on Industrial Engineering

www.journal-aprie.com

J. Appl. Res. Ind. Eng. Vol. 11, No. 4 (2024) 536–559.

Paper Type: Research Paper

Developing a Predictive Model for Engineering Graduates Placement Using a Data-Driven Machine Learning Approach

Md Jakir Hossain Molla[1], Sk Md Obaidullah[1], Soumya Sen[2], Gerhard-Wilhelm Weber[3,4], Chiranjibe Jana[5,*]

[1] Department of Computer Science and Engineering, Aliah University, Kolkata-7000160, India; jakirhossain20@gmail.com; sk.obaidullah@gmail.com.
[2] A.K.Choudhury School of Information Technology, University of Calcutta, Kolkata-700106, India; iamsoumyasen@gmail.com.
[3] Department of Engineering Management, Poznan University of Technology, Poznan, Poland; gerhard.weber@put.poznan.pl.
[4] Institute of Applied Mathematics, METU, Ankara, Turkey; gerhard.weber@put.poznan.pl .
[5] Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai 602105, Tamil Nadu, India; jana.chiranjibe7@gmail.com.

Citation:

Received: 20 July 2024 *Revised: 15 September 2024* *Accepted: 01 October 2024*	Molla, Md., Obaidullah, Sk., Sen, S., Weber, G., & Jana, Ch. (2024). Developing a predictive model for engineering graduates placement using a data-driven machine learning approach. *Journal of applied research on industrial engineering, 11*(4), 536-559.

Abstract

This study presents a novel predictive model for engineering graduates' placement outcomes using Machine Learning (ML) techniques. The model is built on a comprehensive dataset that includes students' performance in various skill areas and their subsequent placement status. By employing a range of ML algorithms, the study evaluates their performance in terms of accuracy. The findings reveal the Customized Random Forest Model (CRFM) algorithm as the most accurate, with a prediction rate of 89%. Furthermore, the study also evaluates the target job domain or field in which students aim to secure placements as well as their target salary packages using the Customized Principal Component Analysis (CPCA) model. The research highlights the importance of various skills, such as programming, aptitude, and domain knowledge, in determining the employability of engineering graduates. The study underscores the importance of various skills, such as programming, aptitude, and domain knowledge, in determining the employability of engineering graduates. The proposed model has directed and practical implications for educational institutions, policymakers, and employers, enabling them to identify the key factors that influence the employability of engineering graduates and develop strategies to enhance their employability.

Keywords: Engineering graduates, Employability, Placement outcomes, Data-driven approach, Predictive analytics, Machine learning.

1 | Introduction

Students pursuing degrees in fundamental sciences, medicine, engineering management, and similar fields often want to get employment in government agencies or big enterprises upon graduation. Numerous pre-

Corresponding Author: jana.chiranjibe7@gmail.com
https://doi.org/10.22105/jarie.2024.468839.1641

placement training programs have been developed and executed throughout the years to enhance the quality and standards of technical education across India. Currently, students' employability is a significant issue for universities, and forecasting their employability in advance might provide early interventions to enhance the institutional placement ratio. The recruiting of students is essential from several perspectives. From a professional perspective, every student must get work. Institutions prioritise placement since it enhances their reputation, thereby elevating their overall ranking and the volume of students they may enroll.

The educational institutions allocate significant resources to provide diverse training programs for its students. Educating students at many stages throughout their studies enhances their industry readiness for employability [1]. Educational institutions often assess potential candidates based on their skill sets. The institutions may use this input to ascertain if the training programs should be sustained or discontinued. Managing a large class size makes individual analysis very difficult.

Rather than depending on human judgment, an analytic model may expedite the process. Diverse evaluations may be conducted for students to determine their current proficiency level. This study may be carried out by using previous data on successful placements to establish a standard for the students. Students can stay in the program until the institution determines they are job-ready.

1.1 | Engineering Graduates

Engineering graduates emerge from their academic journey equipped with a potent blend of theoretical knowledge and practical skills, ready to embark on diverse paths in the world of innovation and problem-solving. Driven by a passion for creating, refining, and implementing solutions to complex challenges, combining innovative thinking, creativity, and technical prowess. Whether they specialize in civil, mechanical, electrical, or any other field, these graduates are the architects of tomorrow's advancements, poised to tackle everything from sustainable infrastructure to cutting-edge technology. Their steadfast dedication to pushing the boundaries of what is possible for engineering graduates is instrumental in shaping the future landscape of industries worldwide.

1.2 | Where Engineering Programs Stand in India Today

According to the Indian Brand Equity Foundation (IBEF), engineering education in India is undergoing fast change as the country works to become a top destination for international students by improving its educational infrastructure and enacting more favourable regulations [2]. While NEP 2020 has the makings of a reform-inspiring policy, its actual implementation is anyone's guess. According to a survey by the Accurate Institute, engineering education in India is not up to par with what is required outside. A lack of competent teachers, poor facilities, and out-of-date curricula that do not equip pupils for the needs of the workforce are all problems [3].

1.3 | Opportunities for Recent Engineering Graduates in the Workforce

Many recent engineering grads, particularly those with degrees in civil engineering and similar disciplines, are finding it difficult to secure entry-level positions in their preferred industries. Reasons for this include a lack of chances, fierce rivalry, and a mismatch between skills and work requirements. While the employment rate in India stayed flat at 46% for four years, the employment ability of engineering graduates rose to 57% in 2023 from 46% in 2021, according to Statista. Factors such as economic conditions, industry demand, and graduation quality have led certain engineering institutions to report reduced campus placements.

Reasons such as greater compensation, more adaptable work descriptions, more room for advancement, and the possibility of self-study influence many engineering students to pursue careers in fields other than engineering. The rigorous academics of engineering school may take a mental and physical toll on students. Mindfulness methods and time management tactics are just two examples of the many internet tools that may help with self-care and effective time management, both of which are crucial for managing stress and avoiding burn out [4].

1.4 | The Unemployment of Recent Engineering Graduates and Their Causes

It is an oversimplified generalization that not all Indian engineers are employable. Engineers graduate in enormous numbers (1.5 million per year) from India; however, their employment varies greatly depending on a variety of factors:

1.4.1 | Factors related to academics

I. Lack of programming & algorithms skills [5].

II. Lack of soft & cognitive skills [6].

III. Lack of domain skills [7].

IV. Lack of english speaking and comprehension skills [8].

V. Have poor analytical and quantitative skills [9].

1.4.2 | Factors related to administration

I. Academic standards at different colleges/university [10].

II. Industry-academia gap [11].

III. Corruption and unethical practices [12].

IV. Economic areas where engineers are not necessary [13].

V. Demand supply mismatch [14].

VI. High student-teacher ratios [15]

VII. Cost of engineering degree [16].

This chart (*Fig. 1*) illustrates the academic-related factors contributing to unemployment among engineering graduates in India, as reported by the national employability report. According to the chart, 91.82% of engineers lack programming and algorithm skills, which is the most significant issue. Additionally, 71.23% lack soft and cognitive skills, while 73.63% have inadequate English speaking and comprehension skills. Other notable issues include 60% lacking domain-specific skills and 57.96% having poor analytical and quantitative skills, all of which hinder their employability.

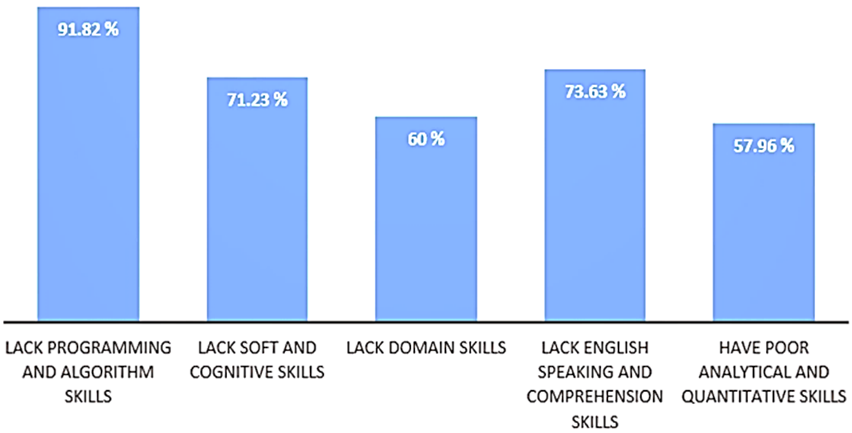

Fig. 1. Unemployment factors related to academics of engineering graduates (source: national employability report).

1.5 | Employability Scenarios of Engineering Graduates in India

The employability gap among engineering graduates has been brought to light by the $245 billion Indian technology sector (Nasscom), which has stressed that the education system is failing to cultivate strong

fundamental and professional abilities. Because of this shortage of qualified workers, IT firms must spend more money and time educating new graduates before they can be deployed. Cost arbitrage and an abundance of engineering degrees have helped India's IT sector flourish for a while, but changing market expectations have led to disappointments.

As the world undergoes a digital revolution, engineers are expected to do more than just fill out paperwork. Unfortunately, abilities such as strong communication, creative thinking, analytical thinking, and problem-solving are frequently overlooked in the current educational system. Industry employers are on the lookout for candidates with strong communication, teamwork, and creative thinking abilities, as well as those with strong backgrounds in AI and cyber security. The major problem is that even after getting a job, many engineering graduates (around 94 per cent) don't have the abilities employers are looking for, and so companies are training them.

According to the aspiring minds research, just 4.77% of participants were able to construct programming logic correctly. Mechanical design engineers and civil engineers are not immune to the current job market; their respective employability percentages are as low as 6.48% and 5.55%. Except for famous institutions like the IITs, the problem of poor engineering education has been known for more than ten years. A worsening of the situation has been the proliferation of subpar engineering schools, several of which have shut down owing to inadequate enrolment.

The AICTE reports that there will be nearly 3.1 lakh fewer engineering seats available over the next four years, including a reduction of about 80,000 seats this year. With dwindling enrolment and no one interested in enrolling, the All India Council for Technical Education (AICTE) has attempted to shutter 800 engineering institutions throughout India. Every year, about 150 colleges choose to close their doors for good. Strict regulations set forth by the AICTE require institutions that do not have adequate facilities and have reported admissions rates below 30% for five years in a row to close their doors [17].

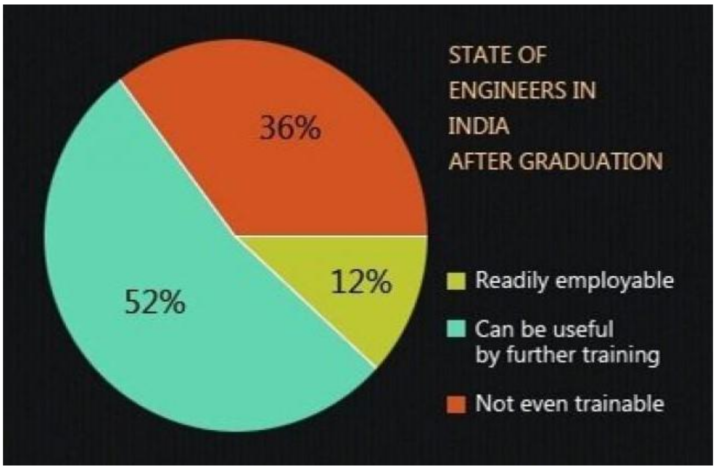

Fig. 2. Placement scenarios of engineers in india (source: Hans India).

1.6 | Justifications of Methodology

To carry out the study, a substantial amount of historical data consisting of students' performance in several skill areas and their employment status is necessary. The situation presented can be classified as a categorization challenge. The prediction will be made by evaluating the various ability levels to see if a student is capable of securing employment [18].

This research study employs several classification techniques, including Support Vector Machine (SVM), Kernel SVM, K Nearest Neighbour (KNN) [19], Decision Tree (DT), and random forest, to assess their performance in terms of accuracy [20]. The dataset utilized in this study comprises information on

engineering/technical students, namely those enrolled in B.Tech programs at various institutions and universities in West Bengal, an eastern state of India. The findings derived from this research can be extrapolated to students in other disciplines as well.

1.7 | Novelties

This research pioneers an innovative approach to addressing the challenge of predicting students' job placements by leveraging a comprehensive data set encompassing their performance across diverse skill sets and their subsequent placement outcomes. By framing the problem as a classification task, the study employs a range of cutting-edge classification algorithms, including logistic regression, DT classifier, SVM, Naïve Bayes, KNN and random forest. this multifaceted analysis not only facilitates a thorough comparison of these methods but also underscores the significance of considering varied skill dimensions in employability assessments.

Moreover, the research's focus on students enrolled in engineering and technical courses in West Bengal, India, adds a novel dimension, enabling insights into the unique employability dynamics within this regional context. Furthermore, the study's assertion regarding the generalizability of its findings to students in other academic disciplines suggests broader applicability, extending the potential impact of the research by a distinct immediate scope. Overall, this research represents a pioneering endeavor at the intersection of educational assessment, predictive analytics, and regional contextualization, offering valuable insights and implications for enhancing students' career prospects across diverse domains. The proposed model depicts the target domain of the students in their placement process. It also predicts their expected salary as per their corresponding domains.

1.8 | Beneficiaries

Researching engineering graduates' placement outcomes can have several potential beneficiaries:

I. Engineering graduates: they can benefit from an improved understanding of the factors that influence their placement outcomes. A predictive model could help them make informed decisions about their career paths, skill development, and job reach strategies.

II. Employers: companies hiring engineering graduates can benefit from a better understanding of the factors that contribute to successful placements. A predictive model could help them identify the most promising candidates and tailor their recruitment processes accordingly.

III. Educational institutions: universities and colleges offering engineering programs can use insights from the research to enhance their curriculum, career counseling services, and alumni support initiatives.

IV. Policy makers: government agencies and policymakers involved in education and workforce development can use the findings to inform policy decisions aimed at improving the employability of engineering graduates.

V. Industry associations: professional organizations representing the engineering sector can use the research findings to advocate for policies and programs that support the professional development and career advancement of engineers. They can also provide support to their members based on the insights gained from the research.

1.9 | Structure of the Paper

The rest of the paper is structured as follows Section 2 describes some related and relevant works on students' placement. Section 3 consists of the research gap. Section 4 sets the objectives of the present work. Section 5 and Section 6 discuss the data and methodology of the work. Section 7 illustrates the results and analysis using tables and graphs. Section 8 designed the proposed model. In Section 9, case studies are conducted with the real dataset. Section 10 depicts the findings of this research. Finally, Section 11 covers the conclusions.

2 | Realted Works

The placement of students is the most important factor for the majority of the universities/colleges worldwide that offer technical/engineering courses. Different types of research work have been carried out in this domain to improve the hit ratio of the institutes.

Ambili [21] used ensemble learning and student database data to predict what characteristics contribute to high-quality learning and campus placement. The recommended model extracts 20 characteristic data. This improved prediction approach based on stacking ensemble learning predicts campus placement options with 90.21% accuracy. Archana et al. [22] used historical student data to predict current student placement rates and suggest ways to improve the institution's placement rate. This technique uses commendations to predict whether a student will be placed. The authors predicted results using the Naive Bayes Classifier and KNN method. They compared both algorithms' performance on the dataset to determine their effectiveness [22].

Davison et al. [23] examined the legality, standardisation, reliability, and job relevance of social media job applicant data. Concerns about this technique were examined. The authors also advise nHR practitioners on using social media for selection goals, preferably with a grasp of caution sweepers. Diana and Priyadharshini [24] evaluated new technologies that firms use for campus recruitment in colleges in 2021. The research also examined how college placement cells affect campus recruiting. The research was descriptive [24]. Jose et al. [25] created a placement predictor that predicts a student's selection for a company based on its criteria. The placement predictor uses numerous factors to assess learner performance.

Algorithm accuracy depends on the task and dataset. Landers and Schmidt have shown that social media is being utilized more and more in the process of selecting and recruiting employees [26]. Parida et al. developed a website-based recommender system that matches applicants' profiles to job requirements. Several Machine Learning (ML) approaches show that the Random Forest Classifier (RFC) has the greatest prediction accuracy. To get the most accurate result, optimisation is used [27]. VidyaShreeram and Muthukumaravel [28] proposed a system to determine if students would continue their study. This is evaluated using ML, a subset of AI. This study uses DT, Random Forest (RF), SVM, and Adaboost to predict students' career prospects [28].

Mohamed et al. [29] incorporated IoT in the recommender system. IoT-based smart commerce systems must discover products and services that will entice consumers to buy. MCDM methods like the MEREC Method were used to weight MABAC criteria for assessing and ranking RSs-IoT [29]. Hezam [30] developed a higher education service quality decision-making model in 2023. The Multi-Criteria Decision-Making (MCDM) approach addresses various evaluation elements. The Promethee method ranks choices. MCDM methods like TOPSIS, VIKOR, EDAS, and MABAC are used for comparative analysis. Metwaly et al. [31] proposed a sophisticated ML system that aims to accurately forecast PM2.5 concentrations in diverse urban settings, taking into account the challenges posed by heterogeneous surroundings.

Abd El-khalik [32] introduced an innovative framework that utilises ML techniques to enhance prediction models for thermal comfort. This paper presents a new method for improving forecasts of thermal comfort by using a large dataset obtained from ASHRAE field experiments and the RP-884 database, which has 107,463 entries. Haratizadeh and Rezaee [33] proposed that ML may be used to develop a novel approach for assessing performance, resulting in a more efficient allocation of capital to portfolio stocks. Nourahmadi and Sadeqi used K-means clustering and Affinity propagation clustering to identify a cohort of investors who had comparable levels of both risk tolerance and risk acceptance.

The authors demonstrate the efficient allocation of assets by using investor characteristics and clustering algorithms [34]. Kuzehgar and Sorourkhah [35] conducted a study to identify the elements that influence student happiness and discontent at a higher education institution. Panda and Shemshad have presented a system that effectively removes unnecessary repetition in human records and simplifies the task of managing attendance [36]. Fakher [37] examined the relationship between environmental sustainability, economic

growth, and FDI in emerging countries. DOLS and FMOLS estimate long-run coefficients. Panel data from 2000 to 2020 are used to evaluate short-term coefficients and causality linkages using Pooled Mean Group (PMG). Long-term predictions show each variable is statistically significant.

Rasinojehdehi and Najafi [38] recommended utilising Data Envelopment Analysis (DEA) to analyse computer network security. Using 10 networks as decision-making units, they used DEA to assess their security. The authors employed Principal Component Analysis (PCA) to reduce inputs and outputs to improve DEA differentiation. Dulal et al. [39] examined COVID-19 data using tensor models. They first extracted patterns across modes using tensor models, canonical polyadic, and higher-order Tucker decompositions.

In addition, they employed canonical polyadic tensor decomposition to anticipate spatiotemporal data from multiple geographical sources and discover COVID-19 hotspots. Xu et al. [40] proposed a structured methodology that included a Subdivision-based Problem Structuring Method (SPSM) for defining the wicked problem, MADM for prioritising sub problems, and DEA for solving one of the most critical subdivisions, bridging the research gap. A case study using the SPSM examines a higher education institution with declining admissions.

Mahboob et al. [41] suggested that the impact of Artificial Intelligence (AI) in education can be viewed as a multi-attribute group decision-making problem in which stakeholders evaluate the pros and cons of AI applications in educational settings based on their preferences and criteria. The methodology's relevance and value are demonstrated by a case study of how AI has affected schooling.

Chen [42] used action research to investigate higher education agency challenges and integrate Online Activity into a university's after-school English program to improve teaching.

Online Activity was chosen as an optimisation technique because it helped teachers adapt online activities to course content, enhancing student interest and outcomes. Bhat [43] noted the neutrosophic AHP group decisionmaking model's weaknesses. An upgraded neutrosophic AHP group decision-making model uses trapezoidal numbers to overcome these constraints.

Shabani et al. [44] examined how Kerman city's government officials' personalities and abilities relate to the universe. Voskoglou and Broumi [45] created a hybrid assessment approach of AR skills in fuzzy conditions using Grey Numbers (GN) and soft sets. An application examines student analogical problem-solving skills. Harouni et al. examined how leadership affects employee communication satisfaction via effective communication. The study found that leader credibility improves communication satisfaction via effective communication [46].

Niksirat [47] examined how virtual education affects student learning. Written and electronic scientific literature were used to gather library data. During the coronavirus pandemic, students attend virtual classrooms. Thus, virtual education's impact on student learning must be studied. The finest education combines online and offline methods.

Hosseinzadeh Kashan et al. [48] created a bi-objective mathematical model for U-shaped assembly line balance that considers setup durations and worker competence and solves it using two methods. Since the research topic is NP-hard, NSGA-II and SPEA-II are utilised to tackle it. The experiments show that NSGA-II outperforms SPEA-II.

Mirsaeidi et al. [49] provided an experimental technique for algorithm selection in forecasting student academic level. After parameter optimisation and algorithm implementation, paired t-tests were used to obtain algorithm performance ratings based on accuracy, f-measure, and ROC. TOPSIS and VIKOR techniques were used to compare algorithms. The suggested technique can pick the optimum educational data mining algorithm.

Deshpande [50] found several supply chain management and production planning courses in industrial engineering degrees. Creative industrial engineering should be included to industrial engineering degrees. This

course will creatively use industrial engineering. Ashbah et al. [51] investigated managers' human qualities, organisation, and performance in a joint stock insurance firm. Field and descriptive-correlational methods were used in this work. The human skills of managers have been studied in relation to organisational performance at Iran Insurance Joint Stock Company, Hamadan branch. Many AI models' intricate inner workings are unknown, sometimes equated to a black box.

Abdullah et al. [52] explored this. Lack of explainability (XAI) decreases responsible AI research, public trust, and desire for openness and interpretability in AI decision-making, especially in critical sectors. This article proposes a Python Extensible Toolkit for Explainable AI (XAI). This toolbox includes nine cutting-edge AI model explanation methods.

3 | Research Gap

It is seldom seen that classifying pupils based on their grades results in improved learning outcomes. This study aims to provide suggestions for employment sectors and placement opportunities for applicants based on their academic performance and many variables, including soft skills and aptitudes, during campus recruiting. The research presented a framework for assigning the appropriate job function and corresponding wage range to applicants during the placement drive. It assists the HR manager in identifying the most suitable applicants according to the company's criteria.

4 | Objectives

Based on the limitations of existing research work, the authors are focused on contributing in the following areas.

I. To identify the essential skill parameters for the campus recruitment process of fresher candidates.

II. To create a model that predicts placement possibilities based on domain knowledge as well as other soft skills.

III. To predict the job domain where a student could be placed based on their set of skills.

IV. To determine the possibility of the salary range of the candidates based on their corresponding job domain.

5 | Data

5.1 | Data Description

The dataset used for this research is student placement data, which has information on 2483 students. The student dataset used in this research contains the information of students who already completed their respective courses. This dataset is based on information collated from the placement departments of several technical/engineering colleges of West Bengal (A state of India) who have passed BE/B. Tech. This is an anonymous dataset, as the names and other student identity information are not given[1].

The dataset contains basic information and a set of attributes, which are described below in *Table 1*.

[1] https://github.com/jakirhossain20/Placement-Data.git

Table 1. Data description.

Variables	Description	Data type/Values
Sex (M/F)	Gender of the candidates	Binary (0: Female, 1: Male)
Branch	Specific department where the student is enrolled (e.g., CSE, EE, ME)	Categorical (B. tech in civil engineering, B.tech in computer science and engineering, b.tech in electronicsand communication engineering, b.tech in electrical engineering, b.tech in mechanical engineering, B. tech in applied electronics and instrumentation, B. tech in information technology)
Score	Evaluation of students in different subjects, including aptitude, english, coding, etc.	Numerical (percentage)
Placement Status	Whether the student has been placed or not	Binary (0: Not Placed, 1: Placed)
Salary	Salary of placed candidates	Numerical (Annual salary in INR)
Job Role	Job role offered to the candidates	Categorical

In the dataset mentioned in Section 5, various parameters were considered to analyze the probability of a fresher candidate's placement based on market research. Most companies conduct their pre-placement exams using the marks from the parameters listed in *Table 1*. However, for the final selection, recruiters may also consider other important attributes, such as co-curricular activities, personality traits, and prior internship experiences.

5.2 | Data Preprocessing

The model has been developed in python 3.7. As Python could not take values, we normalized the dataset. Here, all text values have been converted into numerical ones. Such as all Placed values noted as 1 and not Placed value 0.

6 | Methodology

This study aims to determine whether the subject knowledge has any impact on placement. For that, we have considered six subjects, such as English, Aptitude, Quantitative, Domain, Computer Fundamentals, and Coding, to measure the impact on placement status. Here, the dependent variable is placement status, which is categorical by nature. Here, 1 means placed, and 0 means not placed. On the basis of the predictor variable, we want to create a cut-off score by which we can predict a student placement possibility after knowing the number obtained in each subject. In the context of analyzing student placements, discriminant analysis can be used to predict the probability of a student getting placed (success) based on various predictor variables.

Discriminate analysis is a statistical technique used to classify a set of observations into predefined classes. It does this by determining a linear combination of predictor variables that provides the best discrimination between the groups. The equation provided below is a linear discriminant function [53].

The form of the equation or function is:

$$Z = C + V_1X_1 + V_2X_2 + \cdots + V_nX_n, \tag{1}$$

where Z=discriminate function,

V=the discriminate coefficient or weight for that variable,

X_i= marks obtained in subject i, where i = english, aptitude, quantitative, domain, computer fundamentals, coding, C=Constant.

Here, the model has deployed a Customized Random Forest Model (CRFM) and Customized Principal Component Analysis (CPCA) for the betterment of the result. The basic random forest model depends on three following fundamental mathematical equations which are

$$\mathrm{MSE} = \frac{1}{n}\sum_{1}^{n}(F_i - Y_i)^2. \tag{2}$$

Here, Mean Square Error (MSE) has determined the minimization between the value returned by the model and the actual value of each sample data for N number of sample size.

$$\mathrm{Gini} = 1 - \sum_{i=1}^{n} {F_i}^2. \tag{3}$$

Gini is the index value for each node of the tree, where the P_i is the frequency of the related class interval of the dataset.

$$E = \sum_{i=1}^{n} -F_i \log_2(F_i). \tag{4}$$

Here, E represents entropy, which is the outcome of the probability for each sample space of the dataset.

From the dataset mentioned in Section 5, the different skill sets (english, aptitude, quantitative, domain, computer fundamentals, and coding) have been received through the placement process. The total number of student participants from the different streams was 2483. So, the sample size would be 2483. Here, the dataset mentioned in *Table 1* is heterogeneous, i.e., a combination of text and numerical data. So, the model normalized the whole dataset using various modules of ML methodologies such as numpy, panda, sklearn, etc. The various parameters of skill sets by which a student is getting their placement are defined in the dataset. Therefore, the model has been implemented using supervised learning algorithms such as linear discriminants, random forests, etc. The skill sets are not dependent on each other, and which skill set is more dominant for getting placement is not easily identified. So one more feature, i.e., a customized version of PCA, is added to the model. The fundamental concept of PCA depends on the following mathematical equation.

$$Z = \frac{X-\mu}{\sigma}. \tag{5}$$

Here, μ is the mean of independent features

$$\mu = \{\mu_1, \mu_{2,} \dots\dots\dots, \mu_n\}.$$

σ is the standard deviation of independent features

$$\sigma = \{\sigma_{1,}\sigma_{2,} \dots\dots\dots, \sigma_n\}.$$

$$\mathrm{Cov}\,(x1, x2) = \sum_{1}^{n} \frac{(x1_i - \overline{x1})\,(x2_i - \overline{x2})}{n-1}. \tag{6}$$

The value of covariance can be positive, negative, or zero.

Positive: as the x1 increases, x2 also increases.

Negative: as the x1 increases, x2 also decreases.

Zeros: no direct relation, Let A be a square nXn matrix and X be a non-zero vector for which

$$AX = \lambda X. \tag{7}$$

For some scalar values, λ, which is known as the eigenvalue of matrix A and X is known as the eigenvector of matrix A for the corresponding eigenvalue.

It can also be written as $AX - \lambda X = 0 \Rightarrow (A - \lambda I)X = 0.$

where I is the identity matrix of the same shape as matrix A. And the above conditions will be true only if $(A–\lambda I)$ will be non-invertible (i.e., singular matrix). That means $|A–\lambda I|=0$.

From the above equation, we can find the eigenvalues or lambda, and therefore, the corresponding eigenvector can be found using the equation $AX=\lambda X$.

7 | Results and Analysis

The F-values here provide a sense of the capability of our chosen variables to predict group membership. The Wilks' Lambdas are also significant. So, all the variables are individually capable of predicting group membership.

Table 2. Wilks' lambda of chosen variables.

	Wilks' Lambda	F	df1	df2	Sig.
Aptitude	0.823	534.884	1	2481	0.000
English	0.828	515.759	1	2481	0.000
Quantitative	0.952	126.118	1	2481	0.000
Domain	0.969	80.659	1	2481	0.000
Computer fundamental	0.939	159.770	1	2481	0.000
Coding	0.882	332.805	1	2481	0.000

Lamda is used to test the equality of the group.

The null hypothesis of the Wilks' Lamdatest is H0: No group reparability.

Table3. Wilks' lambda for tests of equality of group means.

Test of Function (s)	Wilks' Lambda	Chi-Square	df	Sig.
1	0.749	717.649	6	0.000

The Wilks' Lambda is statistically significant, as revealed by the statistical significance of the Chi-square statistic. The P-value is less than 0.05. So, we reject the null hypothesis and conclude that the discriminate function is significant.

The discriminate function may be written as

$$Z = -3.58 + 0.43\text{ Aptitude } + 0.025\text{ English} - 0.010\text{ Quantitative} - 0.02\text{ Domain} + 0.009\text{ Computer Fundamental } + 0.020\text{ Coding}. \tag{8}$$

Table 4. Functions at group centroids.

Placement Status	Function
	1
Not placed	-0.418
Placed	0.803

Using the values of group centroids, we can compute the value of Zcrit. It may be used to determine the membership, but it is noting the dataset.

Here Zcrit=0.

Students with dependent variable (Z) values greater than or equal to 0 can now be categorized as placed, whereas students with Z values less than or equal to 0 can be classified as not placed.

Confusion matrix

A confusion matrix is a table used to evaluate the performance of a classification model. It summarizes the predictions against the actual true labels. In this case, we have two confusion matrices: one for the training set and one for the test set.

Table 5. Confusion matrix with training set.

	Predicted	
	Not Placed	**Placed**
Not placed	1167	140
Placed	411	269

The predicted model is shown in *Eq. (2)*. *Table 5* represents the confusion matrix with the training dataset. Here, the accuracy is 72.2%.

Table 6. Confusion matrix with test set.

	Predicted	
	Not Placed	**Placed**
Not placed	290	101
Placed	36	69

Table 6 depicts the confusion matrix with the test dataset with an accuracy of 72.3%.

Predictive measures

AUC=0.930.

Note 1: the cut-off value is set to 0.5.

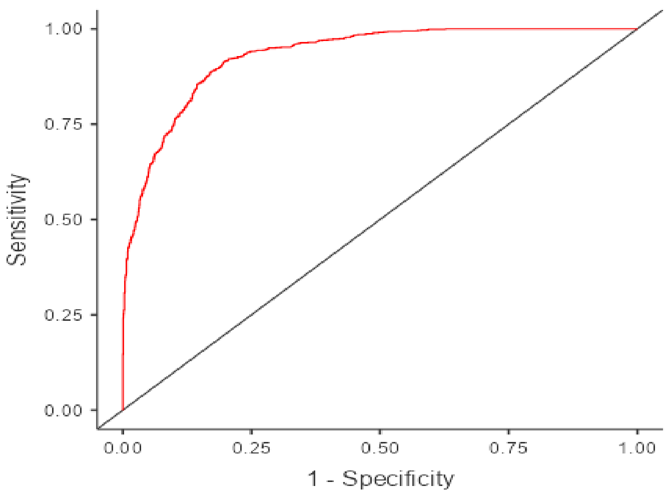

Fig. 3. Rl OC curve.

AUC (Area under the ROC curve)

The AUC is a measure of the model's ability to distinguish between positive and negative classes. In this case, the AUC is 0.930. A higher AUC indicates better separation between classes.

ROC curve

A ROC curve is a graphical representation of the model's performance at different thresholds. It plots the true positive rate (sensitivity) against the false positive rate (1 - Specificity) at various threshold settings. The ROC curve is a comprehensive representation of the model's performance, and the AUC is a single number that summarizes the ROC curve. The link between confusion matrix and AUC/ROC Curve.

To understand the connection, let's analyze the confusion matrices:

I. True Negatives (TN): the number of actual negative instances (not placed) that are correctly predicted as negative. In the training set, TN = 1167, and in the test set, TN = 290.

II. True Positives (TP): the number of actual positive instances (placed) that are correctly predicted as positive. In the training set, TP = 269, and in the test set, TP = 69.

III. False Negatives (FN): the number of actual positive instances that are misclassified as negative. In the training set, FN = 140, and in the test set, FN = 101.

IV. False Positives (FP): the number of actual negative instances that are misclassified as positive. In the training set, FP = 411, and in the test set, FP = 36.

Using these values, we can calculate the true positive rate (sensitivity) and false positive rate (1 - Specificity) at different thresholds. These rates are then plotted on the ROC curve.

The AUC is calculated by integrating the area under the ROC curve. A higher AUC indicates a better separation between classes, which is reflected in the confusion matrices. In this case, the AUC is 0.930, indicating a good separation between the not placed and placed classes.

In summary, the confusion matrices provide a snapshot of the model's performance at a specific threshold (0.5 in this case), while the ROC curve and AUC provide a more comprehensive evaluation of the model's performance across different thresholds.

First, we apply a linear discriminate analysis model to predict the possibility of student placement. This model's accuracy is 72.3%, which is standard but needs improvement. Therefore, we apply ML algorithms, logistic regression, DT classifier, NAÏVE Bayes, KNN, SVM and random forest. random forest enables the college placement cell to input the academic scores of students for predictive insights on placement and sector-specific improvement recommendations.

A predictive model, trained on institutional data, forecasts placement prospects accurately. The proactive engagement between students and authorities, streamlining placements and enhancing employability according to merit. Leveraging modern technology and predictive analytics optimizes efficiency and improves student placement outcomes [54].

We assume the total mark 800 is indexed by 1 as per the Gini Index, as mentioned in *Eq. (3)*. The score of others' marks is treated as Fi.

The higher value of the Gini index is treated as a prominent feature of the dataset, which is determined by random forest application. The dependent variable is categorical placement status. The ML model has passed through 6 algorithms, and the results are displayed below.

Table 7. Comparison accuracy of several ML algorithms.

S.No.	Name of the Algorithms	Without Scale and without Main Features (%)	Without Scale and with Main Features (%)	With Scale and without Main Features (%)	With the Scale and with Main Features (%)
1	Logistic regression	72.0	71.8	72.0	71.8
2	DT classifier	68.8	74.9	68.6	74.5
3	SVM	72.0	73.3	77.9	77.5
4	Naïve bayes	73.9	72.4	73.9	72.4
5	KNN sclassifier	71.2	74.7	66.7	71.0
6	RFC	75.7	78.7	75.7	78.8

From *Table 7* above, it has been found that the RFC has given the highest accuracy, i.e., 78.8 %. Now, the study has designed a novel approach based on the RFC.

8 | Design of the Proposed Models

8.1 | Customized Random Forest Model for Placement Possibilities

From the dataset mentioned under Section 5, where it has been found that 8 features have values 0-100 discretely, the dataset needs some tuning or scaling. For example, if a student has a score of 73 as it is below 75, it is treated as 70, but if the score is equal or more than 75, then it is treated as 80.

Due to this scaling, all the scores lie in the same interval, which is 10. Each algorithm has shown four types of accuracy:

I. Without scale, without main features: the model is considered without any scaling and categorical features.

II. Without scale and with main features: in this category, no scale is required, but only considering significant skill sets for the corresponding category of job.

III. With scale and without main features: here, each mark from the skill set of the student has been scaled up using ceiling and floor functions for all the values without considering any categorical features.

IV. With the scale and main features: here, each mark from the skill set of the student is scaled up using ceiling and floor functions. The model also added the main dominant feature from the student's skill set. For example, analytical or quantitative marks are the most significant factors in getting IT-related jobs. So, this kind of skill set is considered for scaling.

Students with dependent variable (Gini index) values greater than or equal to 0 can now be categorized as placed. In contrast, students with Gini index values less than or equal to 0 can be classified as not placed.

Fig. 4 presents a proposed model where we tune the Random Forest Model to get better output.

Fig. 4. Flow of the proposed model.

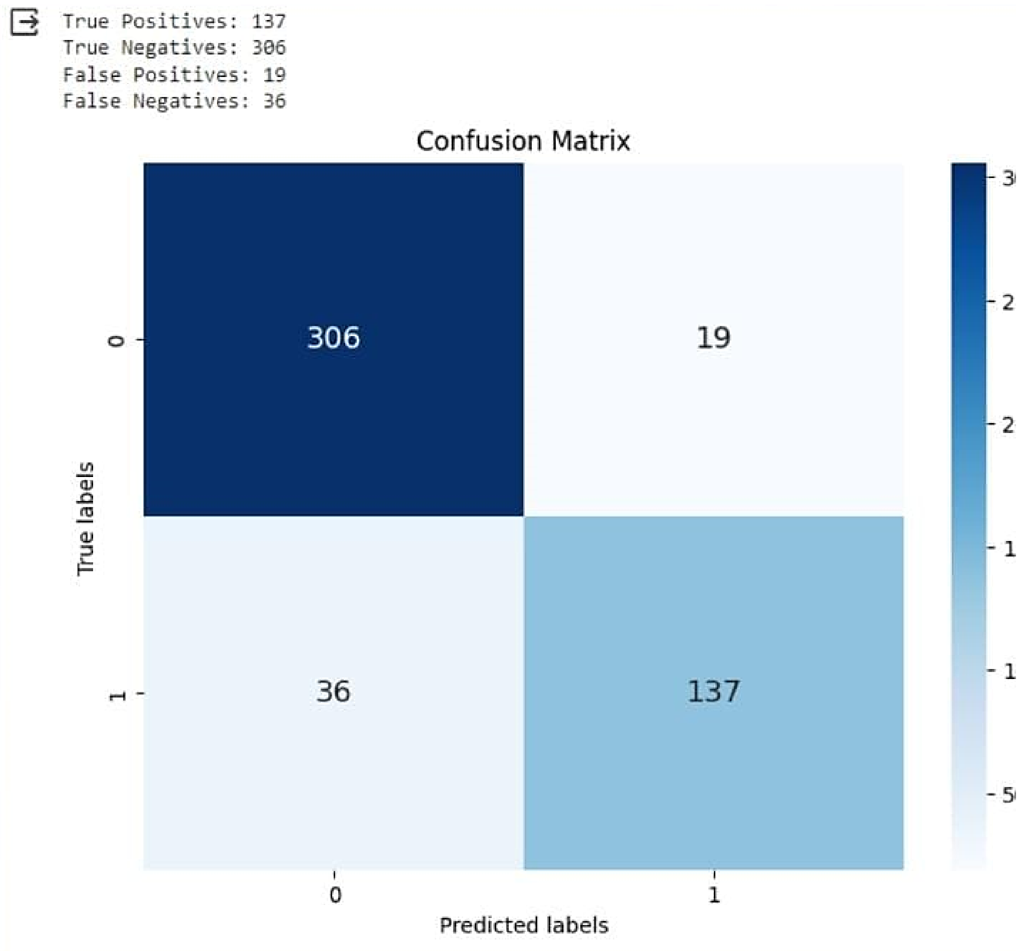

Fig. 5. Confusion matrix with test set.

Fig. 5 shows the confusion matrix with an 88.9% accuracy of the CRFM, which executed the test dataset using Python 3.7 on the Google Colab platform.

Table 8. Classification report.

	Precision	Recall	F1-Score	Support
0	0.89	0.94	0.92	325
1	0.88	0.79	0.83	173
Accuracy			0.89	498
Macro verge	0.89	0.87	0.88	498
Weight age verage	0.89	0.89	0.89	498

Table 8 depicts the CRFM classification reports, which show the benchmark assessment parameters of the CRFM, such as precision, recall, F1-Score, and support.

8.2 | CPCA for Job Domain Selection

Here, we customized the PCA model and formed a CPCA, where we determined the dominant category or skill of the student and its corresponding package. This CPCA has evaluated the most dominant category among 8 discriminant features of a particular student.

After applying the CRFM, it was found that the prediction of classification of placement for a student is high. However, CRFM could not evaluate the proper domain of placement and the corresponding salary range of the students, which was enlightened by Principal Component Analysis (PCA).

PCA increases computing efficiency and can improve ML model performance by concentrating on the most important components. It reduces the dimension of the dataset by which we can identify the most dominant features of the data or criteria of the student, which are identified as skills [55].

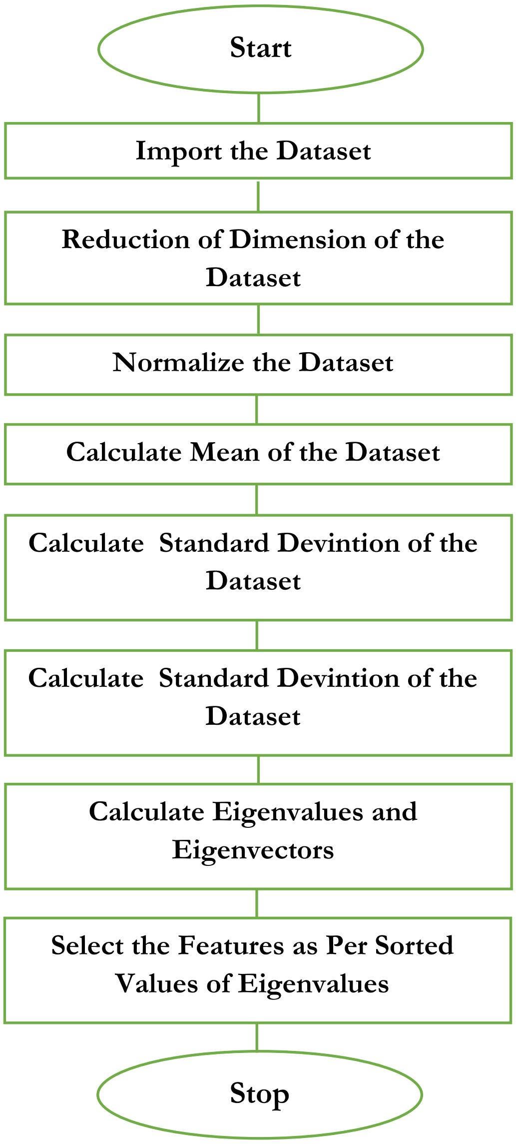

Fig. 6. CPCA flowchart.

The following steps are:

I. First, the dataset has been normalized as per the below *Table 9*.

Table 9.a. Normalize the data set.

Degree	Score/800	Aptitude	English	Quantitative	Analytical	Domain
0	584	69	87	53	67	30
0	584	76	80	73	73	55
0	578	71	73	73	67	55
0	578	71	93	40	80	50
0	575	64	73	47	73	55

Table 9.b. Normalize the data set.

Comp. Funda	Coding	Personality	Placement Status	Salary	JD
20	60	0.0	1	3.5	1
40	0	0.0	0	0	0
47	20	0.0	0	0	0
33	20	0.0	0	0	0
53	20	1.0	0	0	0

II. Start with data: begin with a dataset that has multiple features/dimensions and reduces, as shown in *Table 10*.

Table 10.a. Reduction of the dimension of data.

Degree	Score/800	Aptitude	English	Quantitative
-0.02	87.54	19.18	33.26	12.22
-0.02	87.54	26.18	26.26	32.22
-0.02	81.54	21.18	19.26	32.22
-0.02	81.54	21.18	39.26	-0.78
-0.02	78.54	14.18	19.26	6.22

Table 10.b. Reduction of the dimension of data.

Analytical	Domain	Comp. Funda	Coding	Personality	Placement Status
12.05	-8.16	-17.47	48.32	-6.17	0.66
18.05	16.84	2.53	-11.68	-6.17	-0.34
12.05	16.84	9.53	8.32	-6.17	-0.34
11.84	11.84	-4.47	8.32	-6.17	-0.34
16.84	16.84	15.53	8.32	-6.17	-0.34

III. Standardize: normalize the data to have a mean of 0 and a standard deviation of 1, as shown in *Table 11*.

Table 11.a. Normalize the version of the mean and standard deviation of the dataset.

Degree	Score/800	Aptitude	English	Quantitative
0.04	1.13	0.17	0.21	0.05
1.13	5024.56	909.96	900.01	895.15
0.17	909.96	176.36	170.61	179.30
0.21	900.01	170.61	311.74	84.69
0.05	895.15	179.30	84.69	351.28

Table 11.b. Normalize the version of the mean and standard deviation of the dataset.

Analytical	Domain	Comp. Funda	Coding	Personality	Placement Status
0.26	0.11	0.20	0.38	0.11	-0.00
935.97	386.13	423.56	848.72	-87.36	15.93
179.10	64.38	64.33	118.43	-16.66	2.65
116.16	53.72	64.59	106.08	-22.69	3.47
101.97	82.06	64.39	135.16	-14.34	1.96

IV. Covariance matrix: calculate the covariance matrix to understand how the variables relate to each other.

V. Eigenvalues and eigenvectors: compute the eigenvalues and eigenvectors of the covariance matrix. The eigenvectors are the directions (principal components), and the eigenvalues indicate their magnitude.

Table 12. List of eigenvalues of the eight features.

	0	1	2	3	4	5	6	7	8	9
0	-99.51	21.02	-13.00	10.86	-36.70	10.03	-6.04	7.89	-0.03	-0.02
1	-97.35	-29.24	12.94	8.46	2.54	-9.40	-4.66	13.35	-0.57	0..03
2	-92.70	-5.89	15.44	8.36	4.73	-6.75	-4.94	10.26	0.00	-0.01
3	-91.25	-13.32	-24.20	6.53	-10.12	-11.57	-4.80	11.96	-0.01	-0.03
4	-85.96	-3.56	-8.37	1.34	11.48	-10.42	-3.17	1.45	0.27	-0.06
5	-89.39	-6.25	15.37	4.20	12.24	20.88	-4.38	19.99	-0.57	0.03
6	-81.69	26.92	0.84	-14.88	-27.18	-19.24	-3.73	6.87	0.30	-0.06
7	-70.07	11.59	-6.82	9.13	-6.64	-2.56	-0.82	-0.40	-0.02	-0.04
8	-71.04	-13.07	-15.64	-2.65	-18.75	12.18	-6.42	11.85	-0.01	-0.01
9	-64.59	-28.57	-1.34	8.97	-15.87	-12.75	-0.84	6.75	-0.29	0.01

VI. F. Sort and Select: sort the eigenvalues and their corresponding eigenvectors in descending order. Choose the top 'n' eigenvectors that correspond to the largest eigenvalues.

VII. Transform Data: project the original data onto these 'n' eigenvectors to get the reduced dataset.

```
Eigenvalues:-
 [5.90686523e+03 4.87599311e+02 2.57487273e+02 1.98514063e+02
 1.65030551e+02 1.34983316e+02 4.12665710e+01 2.69440347e+01
 1.68717454e-01 3.87053736e-02]

Eigenvectors:-
 (11, 11)
```

Fig. 7. Best selection of eigenvalues of a student with corresponding eigenvectors.

Here, based on the highest value, we can determine the dominant skills of the students and predict their target categories of placement.

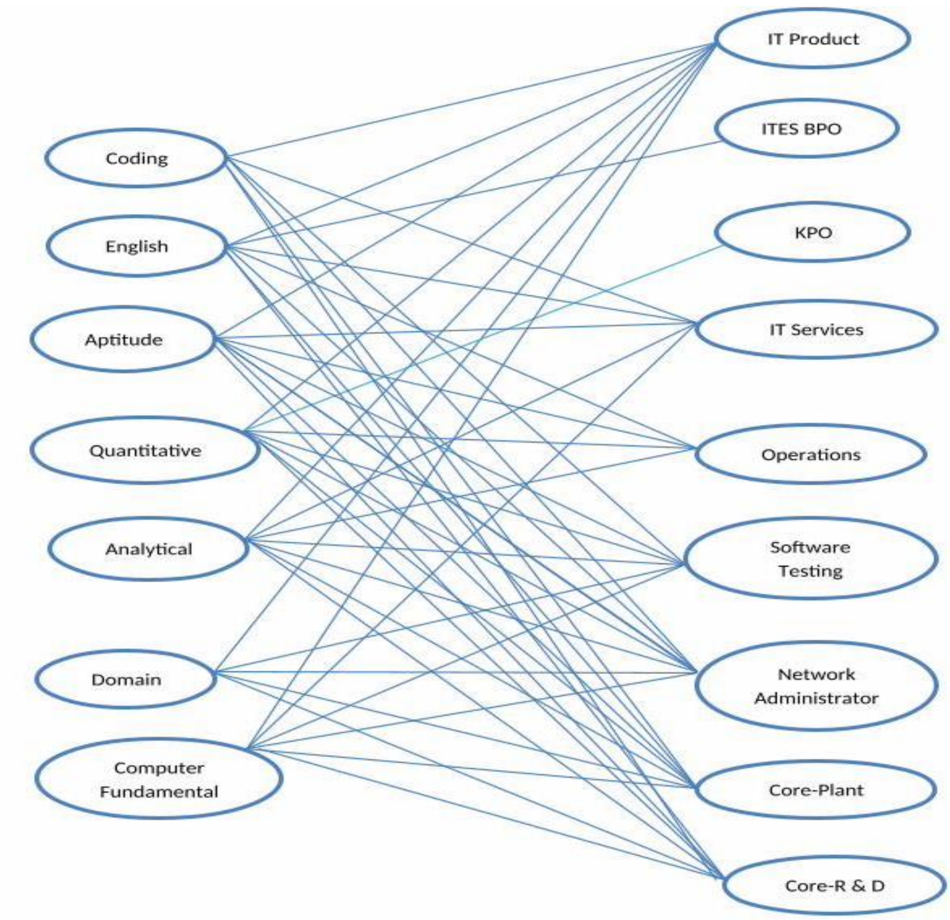

Fig. 8. Mapping diagram between skills set and job domain.

From the above *Fig. 8*, the study depicts the domain selection of the candidates based on their skill set. viz., if a candidate has shown coding, english, aptitude, quantitative, analytical, domain and computer fundamental skills, then he or she will get placement in the domain of IT Product.

9 | Case Study on Real Dataset

In this case study, the simulation tool utilized eight random values (the total score and seven corresponding subjects). It was found that based on their marks in these parameters, candidates could qualify for various job roles offered by the recruitment organization. *Fig. 11* clearly shows that if a candidate achieves the specified marks, they are eligible for the listed job roles except for IT Product roles. To qualify for an IT Product role, additional training programs are required.

Placement Predictor

Total Score: 699

Aptitude: 65

English: 78

Quantitative: 74

Analytical: 64

Domain: 68

Computer Fundamentals: 72

Coding: 79

Submit

Role	Status	Package (Approximately)
IT Product	Need Training	N/A
ITES BPO	Good to go	₹2.41 Lacs
KPO	Good to go	₹3.9 Lacs
IT Services	Good to go	₹5.74 Lacs
Operations	Good to go	₹3.83 Lacs
Software Testing	Good to go	₹4.71 Lacs
Network Administrator	Good to go	₹3.34 Lacs
Sales	Good to go	₹3.12 Lacs
Core Plant	Good to go	₹5.31 Lacs
Core R & D	Good to go	₹6.04 Lacs

← Go Back

Fig. 9. Snapshots of simulation tool on real data.

Fig. 9 also predicts that a student could receive a salary package corresponding to their job roles as depicted in the dataset in Section 5.

10 | Finding

I. In this research, it has been found that the skills or parameters that have been influenced by getting a job in the campus placement process are aptitude, English, quantitative, domain, computer fundamentals, coding, etc.

II. Initially, we employed a linear discriminant analysis model to forecast the likelihood of student placement. The precision of this model is 72.3%, which is considered average but might benefit from improvement. Thus, we utilize ML methods such as logistic regression, DT classifiers, Naïve Bayes, KNN, SVM, and random forests. The random forest model has a relatively high level of accuracy compared to other ML methods. Next, we utilize the random forest model and customize it to produce CRFM, which forecasts the potential outcomes of the placement drive.

III. This study explains the CRFM models, in which we get more accurate values than the base model. In the base model, accuracy was found to be 78.8%, whereas the accuracy for the stated approach was 89%. In this study, we have applied the model to Google Colab, which is a cloud platform. This platform optimizes the computation of algorithms and helps us execute large datasets.

IV. The study has also introduced a novel approach called CPCA to predict the job domain of candidates along with the expected salary range.

11 | Conclusion

To get a proper placement in the industry is very difficult for a fresh candidate. The model depicts the possibility of getting a job using ML algorithms. In conclusion, the CRFM has shown higher accuracy in placement prediction than the linear discriminant model, which has not shown satisfactory results. The CRFM algorithm examines academic performance, skill levels, and personal level data to predict student placements with accuracy.

Students can better grasp their placement odds and improve their techniques with the help of this predictive model. But, issues including students' employability, competency requirements, and market realities need to be addressed. All things considered, this algorithm is a useful tool for improving career chances and placing applicants as effectively as possible. The study has also used the CPCA model to assess the target domain or field for student placements as well as their target compensation packages.

The CRFM models, which yield more accurate results than the base model, are explained in this article. It has been discovered that the accuracy of the base model is 78.8%, whereas the accuracy of the proposed technique is 89%. In addition, a brand-new method known as the study has unveiled CPCA to forecast candidates' predicted wage range and work domain.

11.1 | Limitations

The limitations of this research are

I. The data is not collected from entire institutions across the country.

II. Also, it does not include exhaustive sets of skills or parameters for the recruitment process.

III. This proposed model has not been tested with different data sets.

11.2 | Future Scopes

In the future:

I. The data should be collected from entire institutions across the country.

II. More skills or parameters should be considered in the future.

III. The different data sets can be applied to the proposed model, which will be more generic.

Acknowledgments

The authors acknowledge the staff members of the faculty for their support.

Author Contributaion

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability

The datasets used in this study can be made available upon request from the corresponding author. The datasets are not publicly available due to confidentiality and ethics. The participants only consented to the publication of aggregated data and confidential care of datasets.

Conflicts of Interest

The authors have declared that they have no conflicting of interest regarding the publication of this paper.

References

[1] Casuat, C. D., Sadhiqin Mohd Isira, A., Festijo, E. D., Sarraga Alon, A., Mindoro, J. N., & Susa, J. A. B. (2020). A development of fuzzy logic expert-based recommender system for improving students'Employability. *2020 11th IEEE control and system graduate research colloquium (ICSGRC)* (pp. 59–62). IEEE. DOI: 10.1109/ICSGRC49013.2020.9232543

[2] Subramanian, A. (2019). *The caste of merit: engineering education in India*. Harvard University Press. https://www.hup.harvard.edu/books/9780674987883

[3] Pawan, K. (2020). An empirical study on NEP 2020 [national education policy] with special reference to the future of indian education system and its effects on the stakeholders. *Journal of management engineering and information technology - JMEIT*, *7*(5). DOI:10.5281/zenodo.4159546

[4] Agrawal, A. K., & Harrington-Hurd, S. (2016). Preparing next generation graduates for a global engineering workforce: insights from tomorrow's engineers. *Journal of engineering education transformations*, *29*(4), 5–12. https://sciresol.s3.us-east-2.amazonaws.com/srs-j/jeet/pdf/volume29/issue4/JEET253.pdf

[5] Topalli, D., & Cagiltay, N. E. (2018). Improving programming skills in engineering education through problem-based game projects with Scratch. *Computers & education*, *120*, 64–74. DOI:10.1016/j.compedu.2018.01.011

[6] de Campos, D. B., de Resende, L. M. M., & Fagundes, A. B. (2020). The importance of soft skills for the engineering. *Creative education*, *11*(8), 1504–1520. https://www.scirp.org/journal/paperinformation?paperid=102528

[7] Chan, C. K. Y., Zhao, Y., & Luk, L. Y. Y. (2017). A validated and reliable instrument investigating engineering students' perceptions of competency in generic skills. *Journal of engineering education*, *106*(2), 299–325. https://onlinelibrary.wiley.com/doi/abs/10.1002/jee.20165

[8] Clement, A., & Murugavel, T. (2015). English for employability: a case study of the english language training need analysis for engineering students in India. *English language teaching*, *8*(2), 116–125. https://eric.ed.gov/?id=EJ1075200

[9] Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, qualitative, and mixed research methods in engineering education. *Journal of engineering education*, *98*(1), 53–66. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2009.tb01005.x

[10] Neelaveni, C., & Manimaran, S. (2015). A study on students satsfacton based on qualty standards of accedtaton N hgher educaton. *Educational research and reviews*, *10*(3), 282–289. https://academicjournals.org/journal/ERR/article-full-text/07E850F50081

[11] Sivananda, S., Sathyanarayana, V., & Pati, P. B. (2009). Industry-academia collaboration via internships. *2009 22nd conference on software engineering education and training* (pp. 255–262). IEEE. https://ieeexplore.ieee.org/abstract/document/4812706/

[12] Vedhathiri, T. (2021). Corruptocracy harms high performing faculty in engineering education. *Asia-pacific journal of educational management research*, *6*(1), 2207–5380. DOI:10.21742/AJEMR.2021.6.1.07

[13] Rubtsova, E., Bogolyubova, I., Starodubtseva, G., & Lyubaya, S. (2019). Technological culture of future engineers in the context of modern socio-economic development of the society. *IOP conference series: earth and environmental science*, *315*(2), 22002. DOI:10.1088/1755-1315/315/2/022002

[14] Fernandes, R., Arora, A., Heinz, H., & Asundi, J. (2000). *Supply and demand for software developers in India*. https://www.academia.edu/download/53670738/Supply_and_demand_for_software_developer20170626-26047-1fxfok2.pdf

[15] Dhumal, V. D., & Raju, A. J. (2013). Ratio of students' to total population, teachers and educational institutions in indian higher education–a comparative study (1951-2011). *Indian journal of commerce and management studies*, *4*(3), 36–39. https://ijcms.in/index.php/ijcms/article/view/480

[16] Bertrand, M., Hanna, R., & Mullainathan, S. (2010). Affirmative action in education: evidence from engineering college admissions in India. *Journal of public economics*, *94*(1–2), 16–29. https://www.sciencedirect.com/science/article/pii/S0047272709001431

[17] Blom, A., & Saeki, H. (2011). *Employability and skill set of newly graduated engineers in India*. https://ssrn.com/abstract=1822959

[18] Ikram, A., Su, Q., Fiaz, M., & Khadim, S. (2017). Big data in enterprise management: transformation of traditional recruitment strategy. *2017 IEEE 2nd international conference on big data analysis (ICBDA)* (pp. 414–419). IEEE. https://ieeexplore.ieee.org/abstract/document/8078853/

[19] Giri, A., Bhagavath, M. V. V, Pruthvi, B., & Dubey, N. (2016). A placement prediction system using k-nearest neighbors classifier. *2016 second international conference on cognitive computing and information processing (CCIP)* (pp. 1–4). IEEE.

[20] Vashisht, M. G., & Grover, R. (2019). Employability profiling of engineering students using clustering. *2019 5th international conference on signal processing, computing and control (ISPCC)* (pp. 191–195). IEEE.

[21] Molla, M. J. H., Obaidullah, S. M., Sen, S., Weber, G. W., & Jana, C. (In Press). Developing a predictive model for engineering graduates placement using a data-driven machine learning approach. *Journal of applied research on industrial engineering*. https://www.journal-aprie.com/article_206619_40c95af7a3ceaad8debe55f081b87b45.pdf

[22] Archana, P., Pravallika, D., Priya, P. S., Sushmitha, S., & Amitha, S. (2023). Student placement prediction using machine learning, *10*(1). https://sifisheriessciences.com/index.php/journal/article/view/1235

[23] Davison, H., Bing, M., Kluemper, D., & Roth, P. (2016). Social media as a personnel selection and hiring resource: reservations and recommendations. In *Social media in employee selection and recruitment: theory, practice, and current challenges* (pp. 15–42). Springer. DOI: 10.1007/978-3-319-29989-1_2

[24] Diana, M. H. S., & Priyadharshini, M. (2021). A study on changing trends in campus placement among colleges in Chennai. *International research journal on advanced science hub*, *3*(3S), 1–6.

[25] Jose, I. T., Raju, D., Aniyankunju, J. A., James, J., & Vadakkel, M. T. (2020). Placement prediction using various machine learning models and their efficiency comparison. *International journal of innovative science and researchtec*, *5*(5), 1005–1009. https://www.ijisrt.com/assets/upload/files/IJISRT20MAY509.pdf

[26] Landers, R. N., & Schmidt, G. B. (2016). Social media in employee selection and recruitment: an overview. *Social media in employee selection and recruitment: theory, practice, and current challenges*, 3–11. https://link.springer.com/chapter/10.1007/978-3-319-29989-1_1

[27] Parida, B., KumarPatra, P., & Mohanty, S. (2022). Prediction of recommendations for employment utilizing machine learning procedures and geo-area based recommender framework. *Sustainable operations and computers*, *3*, 83–92. https://www.sciencedirect.com/science/article/pii/S2666412721000489

[28] Sundhararajan, M., & Karthik, B. (2021). *I3CAC 2021: Proceedings of the first international conference on computing, communication and control system, I3CAC 2021*. EAI Publishing. https://eudl.eu/proceedings/I3CAC/2021?articles_page=1

[29] Mohamed, M., Salem, A., & Voskoglou, M. (2024). Attract human loyalty: revealing innovative recommender system using nebulous and intelligent techniques in virtual business realm. *Multicriteria algorithms with applications*, *4*, 69–78. https://sciencesforce.com/index.php/mawa/article/view/331

[30] Hezam, I. M. (2023). A multi-criteria decision making approach for evaluating service quality in higher education. *Multicriteria algorithms with applications*, *1*, 58–66. https://sciencesforce.com/index.php/mawa/article/view/67

[31] Metwaly, A. A., Sleem, A., & Elhenawy, I. (2023). Integrating machine intelligence to estimate PM2. 5 concentration for sustainable urban air quality management. *Sustainable machine intelligence journal*, *4*, 1–6. http://sciencesforce.com/index.php/smij/article/view/48

[32] Abd El-khalik, W. (2022). A machine learning approach for improved thermal comfort prediction in sustainable built environments. *Sustainable machine intelligence journal*, *1*, 1–2. http://sciencesforce.com/index.php/smij/article/view/29

[33] Haratizadeh, S., & Rezaee, F. (2023). A novel machine learning approach for portfolio optimization. *Journal of decisions and operations research*, *8*(2), 527–539. https://www.journal-dmor.ir/article_145630_en.html

[34] Nourahmadi, M., & Sadeqi, H. (2021). A Clustering of investors' behavior according to their financial, behavioral, and demographic characteristics (an application of k-means algorithm). *Innovation management and operational strategies*, *2*(2), 180–194. https://www.journal-imos.ir/article_133336.html?lang=en

[35] Kuzehgar, M., & Sorourkhah, A. (2024). Factors affecting student satisfaction and dissatisfaction in a higher education institute. *Systemic analytics*, *2*(1), 1–13. DOI:10.31181/sa2120247

[36] Panda, A., & Shemshad, A. (2021). Automated class student counting through image processing. *Big data and computing visions*, *1*(1), 24–29. https://www.bidacv.com/article_142509.html

[37] Fakher, H. A. (2021). The role of environmental sustainability, foreign direct investment and trade openness in economic growth: with emphasis on the causal linkage. *Big data and computing visions*, *1*(2), 57–70. https://www.bidacv.com/article_142227_62fac59d4e733d5ae5e6434fc84ce5de.pdf

[38] Rasinojehdehi, R., & Najafi, S. E. (2023). Integrating PCA and DEA techniques for strategic assessment of network security. *Computational algorithms and numerical dimensions*, *2*(1), 23–34. https://www.journal-cand.com/article_183772.html

[39] Dulal, D., Goudarzi Karim, R., & Navasca, C. (2024). Covid-19 pandemic data analysis using tensor methods. *Computational algorithms and numerical dimensions*, *3*(1), 17–44. https://www.journal-cand.com/article_193189.html

[40] Xu, W., Edalatpanah, S. A., & Sorourkhah, A. (2023). Solving the problem of reducing the audiences' favor toward an educational institution by using a combination of hard and soft operations research approaches. *Mathematics*, *11*(18). https://www.mdpi.com/2227-7390/11/18/3815

[41] Mahboob, A., Ullah, Z., Ovais, A., Rasheed, M. W., Edalatpanah, S. A., & Yasin, K. (2024). A MAGDM approach for evaluating the impact of artificial intelligence on education using 2-tuple linguistic q-rung orthopair fuzzy sets and Schweizer-Sklar weighted power average operator. *Frontiers in artificial intelligence*, *7*, 1347626. https://www.frontiersin.org/articles/10.3389/frai.2024.1347626/full

[42] Chen, H.-K. (2023). An action research of student cognative english program: the online activity teaching strategy applications. *Big data and computing visions*, *3*(3), 104–110. https://www.bidacv.com/article_179888.html

[43] Bhat, S. A. (2023). An enhanced AHP group decision-making model employing neutrosophic trapezoidal numbers. *Journal of operational and strategic analytics*, *1*(2), 81–89. https://library.acadlore.com/JOSA/2023/1/2/JOSA_01.02_05.pdf

[44] Shabani, N., Fakheri, S., Fallah, M., Najafi, S. E., & Pour, N. S. (2024). Studying the personality type of employees in governmental organizations of kerman city and its relation to the employees' capability. *Systemic analytics*, *2*(1), 27–34. http://www.sa-journal.org/index.php/systemic/article/view/10

[45] Voskoglou, M., & Broumi, S. (2022). A hybrid method for the assessment of analogical reasoning skills. *Journal of fuzzy extension and applications*, *3*(2), 152–157. https://www.journal-fea.com/article_149915.html

[46] Ghaedamini Harouni, A., Ebrahimzadeh Dastjerdi, R., & Torabi, M. (2022). Investigating the effects of leader credibility on employee communication satisfaction by effective communication. *Innovation management and operational strategies*, *3*(1), 96–112. https://www.journal-imos.ir/article_138388.html?lang=en

[47] Niksirat, M. (2023). Investigating the effect of virtual education on students' learning. *Innovation management and operational strategies*, *4*(1), 63–72. https://www.journal-imos.ir/article_147268_515c78a7386007892f218663ea1cf632.pdf

[48] Kashan, A. H., Afkhami, S., & Maroofkhani, P. (2023). Multi-objective optimization of the u-shaped assembly line balancing and worker assignment considering setup-times and workers' skill. *Journal of decisions & operations research*, *8*(2), 283-306. https://www.journal-dmor.ir/article_136827_1c91b66e674e40be4a20d6f429cf75de.pdf

[49] Mirsaeedi, F., Koosha, H., & Ghodoosi, M. (2021). Evaluation of data mining algorithms on educational data using multi-criteria decision-making methods. *Journal of decisions and operations research*, *6*(1), 41–55. https://www.journal-dmor.ir/article_129587.html

[50] Deshpande, P. V. (2023). New course in industrial engineering education curriculum. *International journal of research in industrial engineering*, *12*(1), 1–6. https://www.riejournal.com/article_166510_32c64eeb750894592be49c9a5702559d.pdf

[51] Ashbah, M., Ahmadi, S. A. A., & Sarlak, M. A. (2022). The effect of managers' human skills on organizational commitment and employee performance of Iran insurance company. *Modern research in performance evaluation*, *1*(3), 202–212. https://www.journal-mrpe.ir/article_171106_aff1194b10c9dcd021b0b4b56dfef033.pdf

[52] Abdullah, W., Tolba, A., Elmasry, A., & Mostafa, N. N. (2024). Visioncam: a comprehensive XAI toolkit for interpreting image-based deep learning models. *Sustainable machine intelligence journal*, *8*, 4–46. https://sciencesforce.com/index.php/smij/article/view/290

[53] Zhang, P., Peng, J., & Riedel, N. (2005). Discriminant analysis: A unified approach. *Fifth IEEE international conference on data mining (ICDM'05)* (pp. 8-pp). IEEE. DOI: 10.1109/ICDM.2005.51

[54] Zermane, A., Mohd Tohir, M. Z., Zermane, H., Baharudin, M. R., & Mohamed Yusoff, H. (2023). Predicting fatal fall from heights accidents using random forest classification machine learning model. *Safety science*, *159*, 106023. https://www.sciencedirect.com/science/article/pii/S0925753522003629

[55] Elsamanty, M., Ibrahim, A., & Saady Salman, W. (2023). Principal component analysis approach for detecting faults in rotary machines based on vibrational and electrical fused data. *Mechanical systems and signal processing*, *200*, 110559. https://www.sciencedirect.com/science/article/pii/S0925753522003629